Workload Characterization

CS 700

Objective

- To observe the key characteristics of a workload, and develop a workload model that can be used to test multiple alternatives
  - Both analytical models and simulations require a workload model
- Example: modeling a web server
  - Inter-arrival process, service demands
    - Need information about distributions, not just summary statistics
  - Classes of requests

Fitting Distributions to Data

- First step: hypothesizing what family of distributions, e.g. Poisson, normal, is appropriate without worrying yet about the specific parameters for the distribution
  - Have to consider the shape of the distribution

3

Heuristics for hypothesizing a distribution

- Summary statistics can provide some information
  - Coefficient of variation (CV)
    - CV = 1 for exponential distribution, CV > 1 for hyper-exponential, CV < 1 for hypo-exponential, erlang
    - But CV not useful for all distributions, e.g., $N(0,\sigma^2)$
  - For discrete distributions, Lexis ratio $\tau = \sigma^2/\mu$ has the same role that CV does for continuous distributions
    - $\tau = 1$ for Poisson, $\tau < 1$ for binomial, $\tau > 1$ for negative binomial

4

Heuristics cont'd

- Histograms
  - Break up the data into k disjoint adjacent intervals of the same width and compute the proportion of data points that lie in each interval
    - Sturge's rule of thumb

      Given n data points $k = \lfloor 1 + \log_2 n \rfloor$
  - Visually compare the shape of the histogram to that of known distributions

5

Estimation of Parameters

- After hypothesizing a distribution, next step is to specify their parameters so that we can have a completely specified distribution
- Several techniques have been developed
  - Method of moments, Maximum likelihood estimators, Least-squares estimators

6

Method of moments

- Compute the first k moments of the sample data
- Equate the first few population moments with the corresponding sample moments to obtain as many equations as there are unknown parameters
  - Solve these equations simultaneously to obtain the required estimates

7

Maximum Likelihood Estimation

- Suppose we have hypothesized a discrete distribution for our data that has one unknown parameter θ. Let $p_\theta(x)$ denote the pmf for this distribution. If we have observed the data $X_1, X_2, \ldots, X_n$, we define the likelihood function $L(\theta)$ as follows:

$$L(\theta) = p_\theta(X_1)p_\theta(X_2)\ldots.p_\theta(X_n)$$

The MLE of θ is defined to be the value of θ that maximizes $L(\theta)$

For continuous distributions, $L(\theta)$ is defined analogously

8

MLE for exponential distribution

$$p(\beta) = 1/\beta e^{-x/\beta}$$

$$L(\beta) = (1/\beta e^{-X_1/\beta})(1/\beta e^{-X_2/\beta})....(1/\beta e^{-X_n/\beta})$$

$$= \beta^{-n} \exp(-\frac{1}{\beta}\sum_{i=1}^{n} X_i)$$

Taking logs on both sides, we have

$$\ln L(\beta) = -n \ln \beta - \frac{1}{\beta}\sum_{i=1}^{n} X_i$$

It can be shown through standard differential calculus by setting the derivative to 0 and solving for β that the value of β that maximizes L(β) is given by

$$\beta = (\sum_{i=1}^{n} X_i)/n = \overline{X}(n)$$

9

Determining how representative the fitted distributions are

- Both heuristic procedures and statistical techniques can be used for this
- Heuristics (Graphical/Visual techniques)
  - Density/Histogram Overplots and Frequency Comparisons
  - Q-Q plots
  - Probability plots (P-P plots)
  - Distribution Function Difference Plots

10

Statistical techniques

- Goodness-of-fit tests
  - Chi-square tests
  - Kolmogorov-Smirnov (KS) tests
  - Anderson-Darling (AD) tests
  - Poisson-process test

11

Chi-square tests

- First divide the entire range of the fitted distribution into k adjacent intervals
- Tally the number of data points in each interval o_i
- Compute the expected proportion of data points in each interval e_i
- Compute $D = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i}$
  - D has a chi-square distribution with k-1 degrees of freedom
  - If the computed D less than $\chi^2(1-\alpha,k-1)$ then the observations come from the specified distribution

12

Chi-square tests cont'd

- Cell sizes should be chosen so that the expected probabilities e_i are all equal
- If the parameters of the hypothesized distribution are estimated from the sample then the degrees of freedom for the chi-square statistic should be reduced to k-r-1, where r is the number of estimated parameters
- For continuous distributions and for small sample sizes, the chi-square test is an approximation

13

Other tests

- Kolmogorov-Smirnov
  - Based on the observation that the difference between the observed CDF $F_o(x)$ and the expected CDF $F_e(x)$ should be small
- Anderson-Darling
  - More powerful in detecting differences in the tails of distributions

14

Fitting distributions to data in practice

- Use distribution-fitting software!
  - ExpertFit software from Averill Law
  - BestFit software
  - Download software and try it out on random data that you generate or data in exercises
    - See links on class web site

15

Clustering

- Many workloads consist of multiple classes of customers/requests
- Clustering is a technique used for classifying requests into multiple groups where members of one group are as "similar" as possible
  - Intragroup variance should be as small as possible whereas intergroup variance should be as large as possible
  - Non-hierarchical clustering: start with k clusters, move members around until intragroup variance is minimized
  - Hierarchical clustering: agglomerative and divisive

16

Minimum spanning tree method

- Agglomerative hierarchical clustering technique
- Algorithm
  1. Start with k = n cluster
  2. Find the centroid of the ith cluster. The centroid has parameter values equal to the average of all points in the cluster
  3. Compute the intercluster distance matrix (distance between centroids)
  4. Find the smallest nonzero element of the distance matrix. Merge the two clusters with the smallest distance and any other clusters with the same distance
  5. Repeat steps 2 to 4 until all components are in the same cluster

Minimum spanning tree method cont'd

- Results of the clustering process can be represented as a spanning tree (a dendrogram) where each branch of the tree represents a cluster and is drawn at a height where the cluster merges with the neighboring cluster
- Given any maximum allowable intracluster distance, by drawing a horizontal line at the specified height we can obtain the desired clusters

Example

Consider a workload with five components and two parameters

Program	CPU time	Disk I/O
A	2	4
B	3	5
C	1	6
D	4	3
E	5	2

19

Example cont'd

First iteration:

	A	B	C	D	E
A	0	$2^{0.5}$	$5^{0.5}$	$5^{0.5}$	$13^{0.5}$
B		0	$5^{0.5}$	$5^{0.5}$	$13^{0.5}$
C			0	$18^{0.5}$	$32^{0.5}$
D				0	$2^{0.5}$
E					0

Minimum inter-cluster distance is between A and B, and D and E. The two pairs are merged

20

Example cont'd

- Second iteration:
  - Centroid of AB is {(2+3)/2),(4+5/2)}, I.e. {2.5,4.5}. Similarly for DE it is {4.5,2.5}

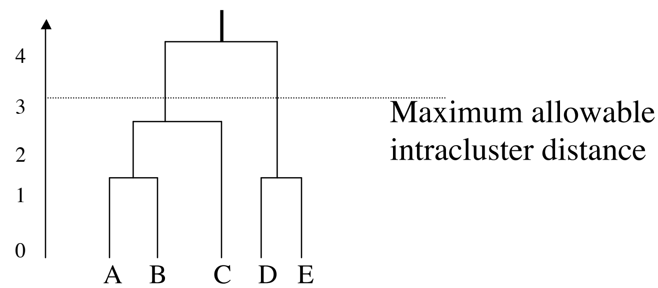
	AB	C	DE
AB	0	$4.5^{0.5}$	$8^{0.5}$
C		0	$24.5^{0.5}$
DE			0

  - Merge AB and C

21

Example cont'd

- Third iteration: merge ABC and DE to get a single cluster ABCDE



22

Additional Reading

- Articles on workload characterization by Calzorossa and Feitelson
  - On class web site